

# REPORT



과목명		실험계획법
담당교수		최영훈 교수님
학과		응용통계학과
학년		3학년
학번		201452024
이름		박상희
제출일		기말고사 족보풀이

# 2017년 실험계획법 기말고사

[문제1] 아래의 모집단 모형을 수식으로 간단히 표현하여라.

– Latin Square Design

$$y_{ijk} = \mu + \alpha_i + \beta_j + \Gamma_k + \epsilon_{ijk} \quad i = 1, \dots, p \quad j = 1, \dots, p \quad k = 1, \dots, p$$

단,  $\mu$ : 모집단 총평균,  $\alpha_i$ : 열(블록)의  $i$ 번째 수준의 효과이며  $\sum_{i=1}^p \alpha_i = 0$ ,  $\beta_j$ : 행(블록)의  $j$ 번째 수준의 효과이며  $\sum_{j=1}^p \beta_j = 0$

$\Gamma_k$ : 처리(요인)의  $k$ 번째 수준의 효과이며  $\sum_{k=1}^p \Gamma_k = 0$   $\epsilon_{ijk} \sim N(0, \sigma^2)$ 이며 독립인 오차항

– 2단계 분지모형

$$y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \epsilon_{k(ij)} \quad i = 1, \dots, a \quad j = 1, \dots, b \quad k = 1, \dots, n$$

단,  $\mu$ : 모집단 총평균,  $\alpha_i$ : 인자  $A$ 의 효과이며  $\sum_{i=1}^a \alpha_i = 0$ ,

$\beta_{j(i)}$ : 인자  $B(A)$ 의 효과이며  $\sum_{j=1}^b \beta_{j(i)} = 0$ ,  $\epsilon_{k(ij)} \sim N(0, \sigma^2)$ 이며 독립인 오차항

– 3단계 분지모형

$$y_{ijkl} = \mu + \alpha_i + \beta_{j(i)} + \gamma_{k(ij)} + \epsilon_{l(ijk)} \quad i = 1, \dots, a \quad j = 1, \dots, b \quad k = 1, \dots, c \quad l = 1, \dots, n$$

단,  $\mu$ : 모집단 총평균,  $\alpha_i$ : 인자  $A$ 의 고정효과이며  $\sum_{i=1}^a \alpha_i = 0$ ,  $\beta_{j(i)}$ : 인자  $B(A)$ 의 효과이며  $\sum_{j=1}^b \beta_{j(i)} = 0$

$\gamma_{k(ij)}$ : 인자  $C(AB)$ 의 효과이며  $\sum_{k=1}^c \gamma_{k(ij)} = 0$ ,  $\epsilon_{l(ijk)} \sim N(0, \sigma^2)$ 이며 독립인 오차항

– 불완전블록모형 (BIB)

$$y_{ijk} = \mu + \alpha_i + \beta_j + \epsilon_{ijk} \quad i = 1, \dots, a \quad j = 1, \dots, b$$

단,  $\mu$ : 모집단 총평균,  $\alpha_i$ :  $i$ 번째 처리수준의 효과이며  $\sum_{i=1}^a \alpha_i = 0$ ,  $\beta_j$ :  $j$ 번째 블록수준의 효과이며  $\sum_{j=1}^b \beta_j = 0$

$\epsilon_{ijk} \sim N(0, \sigma^2)$ 이며 독립인 오차항

– 공분산분석모형 (한 개의 공변수가 존재하는 경우)

$$y_{ij} = \mu + \tau_i + \beta(x_{ij} - \overline{x_{\cdot \cdot}}) + \epsilon_{ij} \quad i = 1, \dots, a \quad j = 1, \dots, b$$

단,  $\mu$ : 모집단 총평균,  $\tau_i$ :  $i$ 번째 처리수준의 효과이며  $\sum_{i=1}^a \tau_i = 0$   $x_{ij}$ : 공변수의 관측값(공변량)

$\epsilon_{ij} \sim N(0, \sigma^2)$ 이며 독립인 오차항

[문제2] 아래의 문항에 대하여 간단히 답하여라.

- EMS column이란 무엇이며, 왜 필요한가?

각 요인들의 효과유무를 검정하기 위한 F 통계량 형태(특히, 분모의 평균제곱합)를 결정하는데 상당히 유효하기 때문이다.

- PROC VARCOMP 프로시저를 적용하는 경우는?

모든 요인이 랜덤인 랜덤효과와 분류변수를 위한 명령문으로 EMS 뿐만 아니라 분산요소들을 추정한다.

- PROC GLM의 RANDOM 분류변수 / TEST; 명령문과 TEST H= E= ; 명령문의 차이는?

PROC GLM의 RANDOM 분류변수 / TEST; 명령문 : EMS를 생성하고 각 요인에 따른 잔차항을 자동으로 파악하여 가설 검정을 수행하는 역할을 한다. 주요인이 랜덤이면, 이 주요인과 결합형태의 상호작용 요인은 랜덤이다.

TEST H= E= ; 명령문 : 이 명령문은 통계실험자가 원하는 가설요인과 이에 따른 가설요인의 잔차항을 직접 지정하는 방법이다.

- 혼합효과란?

두 요인 가운데 A 요인은 고정인 B 요인은 랜덤인 혼합 효과 모형의 분석을 혼합 모형 분산 분석이라고 한다.

- 라틴방격모형의 특징은?

- (1) P개의 처리수준이 존재하며, 알파벳(A, B, C, D)으로 표시한다.
- (2) 각각 P 수준인 두 종류의 블록변수가 존재한다.
- (3) 각각의 행과 열은 모든 처리수준을 포함한다.
- (4) 일반적으로 각 효과의 수준이  $3 \leq P \leq 7$  인 실험에 적용된다.

- 분지(Nested)모형의 특징은?

분지모형은 한 요인의 수준이 다른 요인의 수준 안에 포함되는 분지 모형으로 요인실험모형을 이용하여 분석할 수 있으며, 랜덤화에 제약을 갖는다.

- B가 A에 포함된 2단계 분지모형의  $SS_{B(A)}$ 를 이원배치법의 제곱합으로 표현하여라.

$$SS_{B(A)} = SS_B + SS_{AB}$$

- 위의 문항에서 B(A)의 자유도는? (단 각 A,B 요인의 수준수를 a, b라 가정)

인자 B(A)의 자유도는 요인 모형의 요인B와 상호작용 AB 자유도의 합과 같다.  $a(b-1) = (b-1) + (a-1)(b-1)$

- 3단계 Nested 모형에서 A B C 요인 (A : 상위단계)에 따른 SAS 명령문은?

MODEL Y = A B(A) C(BA);

- BIB 모형의 특징은?

모든 처리 수준이 매 블록마다 적용될 수 없는 실험계획모형이다.

- 공분산분석(ANCOVA)이란?

분산분석과 회귀분석의 특성을 결합한 분석기법 => 공변수의 효과를 보정한 분산분석의 기법

- 공분산분석에서 분류변수가 store, 종속변수가 y, 공변수가 x 일 때의 PROC GLM의 model 명령문은?

model = time = store price / SOLUTION;

- 위 문항에서 PROC GLM의 model 명령문 옵션으로 /SOLUTION; 의 용도는?

모수의 추정량  $\hat{\tau}_i$  및 최소제곱추정량  $\hat{\beta}$ 을 출력한다.

[문제3] 2요인 실험계획 (단 A 요인이 고정, B 요인이 랜덤) 하에서 모집단 모형이

$$Y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \epsilon_{ijk}, \quad i = 1,2,3 \quad j = 1,2,3,4 \quad k = 1,2 \text{ 일 때}$$

- EMS column 을 완성하여라.
- $\alpha_i$   $\beta_j$   $\alpha\beta_{ij}$  효과 유무 검정을 위한 가설과 F 통계량은?

요인	3 F i	4 R j	2 R k	F 통계량
$\alpha_i$	0	4	2	$MS_A / MS_{AB}$
$\beta_j$	3	1	2	$MS_B / MSE$
$\alpha\beta_{ij}$	0	1	2	$MS_{AB} / MSE$
$\epsilon_{(ij)k}$	1	1	1	

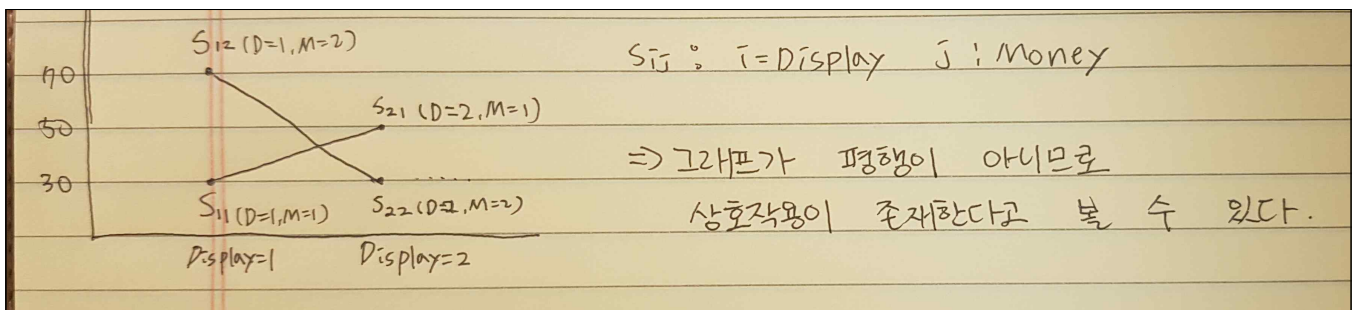
[문제4] 광고문안(Display)과 광고비(Money)가 기업의 판매액(sales)에 미치는 영향을 조사하고자 3회 반복이 존재하는 이원배치법(2x2 요인계획법)을 계획하였다. 이 때 광고 문안은 고정(Fixed)이며, 광고비는 랜덤(Random)으로 추출하였다.

Display \ Money	1	2
1	10 30 50	40 50 60
2	60 70 80	10 30 50

- DO loop 를 이용한 SAS 프로그램을 완성하여라. (단 문제3의 EMS 결과 참조)

```
DATA;
DO Display = 1 to 2;
DO Money = 1 to 2;
DO t = 1 to 3;
INPUT sales @@;
OUTPUT;
END; END; END;
CARDS;
10 30 50 40 50 60
60 70 80 10 30 50
;
PROC GLM;
CLASS Display Money;
MODEL sales = Display Money Display*Money;
RANDOM Money Display*Money / TEST;
TEST H=Display Money E=Display*Money;
MEANS Display / SNK E=Display*Money;
```

- interaction(상호작용)이 존재하는지를 파악하기 위한 profile plot (측면도 그래프)를 수작업으로 그리고, 상호작용 여부를 결정하여라.



[문제5] 아래 결과를 이용하여 문항에 답하여라.

Source	DF	Type III SS	Mean Square	F Value	Pr > F
area	1	315.375000	315.375000	1.77	0.2084
state(area)	4	6453.833333	1613.458333	9.04	0.0013
city(area*state)	6	2226.250000	371.041667	2.08	0.1322

Tests of Hypotheses Using the Type III MS for (1) state (area\*state) as an Error Term

Source	DF	Type III SS	Mean Square	F Value	Pr > F
area	1	315.375000	315.375000	0.85	0.3921
state(area)	4	6453.833333	1613.458333	4.35	0.0545

Student-Newman-Keuls Test for opinion

SNK Grouping	Mean	N	area
A	56.667	12	1
A	49.417	12	2

---

Variance Components Estimation Procedure

Source	DF	SumSquares	Mean Square	Expected Mean Square
area	1	315.375000	315.375000	Var(Error) + 2 Var(city(area*state)) + Q((area, state(area)))
state(area)	4	6453.833333	1613.458333	Var(Error) + 2 Var(city(area*state)) + Q(state(area))
city(area*state)	6	2226.250000	371.041667	Var(Error) + 2 Var(city(area*state))
Error	12	2141.500000	178.458333	Var(Error)
Total	23	11137.000000		

Type 1 Estimates

Variance Component	Estimate
Var(city(area*state))	(2)
Var(Error)	(3)

---

(The NESTED) Procedure

Source	area	state	city	Error
area	12	4	2	1
state	0	4	2	1
city	0	0	2	1
Error	0	0	0	1

Nested (4) Effects Analysis of Variance for Variable opinion

Variance Source	DF	Sum of Squares	F Value	Pr > F	Error Term	Mean Square	Variance Component	Percent of Total
Total	23	11137				484.215580	585.354167	100.0000
area	1	315.375000	0.20	0.6813	state	315.375000	-108.173611	0.0000
state	4	6453.833333	4.35	0.0545	city	1613.458333	310.604167	53.0626
city	6	2226.250000	2.08	0.1322	Error	371.041667	96.291667	16.4502
Error	12	2141.500000				178.458333	178.458333	30.4872

- 실험계획 모형을 밝혀라.

$$y_{ijk} = \mu + \alpha_i + \beta_{j(i)} + \gamma_{k(ij)} + \epsilon_{l(ijk)} \quad i = 1, \dots, a \quad j = 1, \dots, b \quad k = 1, \dots, c \quad l = 1, \dots, n$$

단,  $\mu$ : 모집단 총평균,  $\alpha_i$ : 인자 A의 고정효과이며  $\sum_{i=1}^a \alpha_i = 0$ ,  $\beta_{j(i)}$ : 인자 B(A)의 효과이며  $\sum_{j=1}^b \beta_{j(i)} = 0$

$\gamma_{k(ij)}$ : 인자 C(AB)의 효과이며  $\sum_{k=1}^c \gamma_{k(ij)} = 0$ ,  $\epsilon_{l(ijk)} \sim N(0, \sigma^2)$ 이며 독립인 오차항

- area, state, city 요인의 Fixed, Random 여부를 결정하고, 이유를 제시하여라.

분산요소의 오른쪽 마지막의 제곱항(상수)이 Q(area, state(area)) 존재하므로 area는 Fixed 요인이다.

분산요소의 오른쪽 마지막의 제곱항(상수)이 Q(state(area)) 존재하므로 state는 Fixed 요인이다.

분산요소의 오른쪽 마지막의 제곱항(상수)이 존재하지 않으므로 city는 Random 요인이다.

- area, state, city 요인효과가 존재하는지를 p-값을 제시한 후 언급하여라.

유의수준 5% 하에서 area의 효과는 p-값이 0.3921로 유의수준 5% 하에서 "area의 요인효과가 존재하지 않는다."라는 귀무가설을 채택한다. 따라서 area의 요인효과는 존재하지 않는다.

유의수준 5% 하에서 state의 효과는 p-값이 0.0545로 유의수준 5% 하에서 "state의 요인효과가 존재하지 않는다."라는 귀무가설을 채택한다. 따라서 state의 요인효과는 존재하지 않는다.

유의수준 5% 하에서 city의 효과는 p-값이 0.1332로 유의수준 5% 하에서 "city의 요인효과가 존재하지 않는다."라는 귀무가설을 채택한다. 따라서 city의 요인효과는 존재하지 않는다.

- 빈칸 (1), (2) (3) (4)를 완성하여라. (단 (4)는 고정, 랜덤, 혼합 중 선택)

(1) city(area\*state)  
(2) 96.29167  
(3) 178.45833  
(4) Random

- area 요인의 수준간의 차이가 존재하는가?

area1은 평균이 56.667, area2는 평균이 49.417 이지만 SNK 검정 결과 차이가 없다.

- area 요인의 p-값 0.6813 의미(적용상의 경우포함)를 해석하여라.

The NESTED Procedure는 모든 효과가 랜덤일 때 사용한다. 따라서 랜덤 효과를 가정한 경우, p-값이 0.6813으로 유의수준 5%하에서 차이가 존재하지 않는다는 귀무가설을 채택한다. 따라서 area 간의 차이가 존재하지 않는다.

- 하나의 요인효과가 존재한다면, Variance Component 항목 값 제시와 함께 밝혀라.  
(하단부 NESTED procedure 결과참조)

state의 분산요소가 310.6042로 총 분산 585.3541 중 53.06 % 으로 가장 많은 비중을 차지한다. 따라서 하나의 요인효과가 존재한다면 state 효과라고 볼 수 있다.

## 2016년 실험계획법 기말고사

[문제1] 아래의 EMS 결과를 이용하여 답하여라.

Source	Type III Expected mean Square
area	$\text{Var}(\text{Error}) + 2 \text{Var}(\text{city}(\text{area} * \text{state})) + Q(\text{area}, \text{state}(\text{area}))$
state(area)	$\text{Var}(\text{Error}) + 2 \text{Var}(\text{city}(\text{area} * \text{state})) + Q(\text{state}(\text{area}))$
city(area+state)	$\text{Var}(\text{Error}) + 2 \text{Var}(\text{city}(\text{area} * \text{state}))$

– PROC GLM; 명령문을 완성하여라.

```
CLASS area state city;
MODEL opinion = area state(area) city(area state);
RANDOM city(area state);
TEST H = area state(area) E = city(state area);
MEANS area / LSD E = city(state area);
```

– 아울러 위문항의 명령문을 아래 명령문으로 완성하여라.

```
PROC VARCOMP;
CLASS area state city;
MODEL opinion = area state(area) city(area state);
```

[문제2] 아래 빈칸의 분산값을 추정하여라.

Source	DF	Sum Squares	Mean Squares	Expected Mean Square
machine	3	89.2	29.7	$\text{Var}(\text{Error}) + 4 \text{Var}(\text{machine})$
Error	12	22.75	1.9	$\text{Var}(\text{Error})$
Total	15	111.95		

Estimate	
Variance Component	Estimate
Var(machine)	
Var(Error)	

```
Var(Error) = 1.9
29.7 = 1.9 + 4Var(machine)
Var(machine) = 6.95
```

[문제3] A 요인은 랜덤(R), B 요인은 고정(F)인 혼합모형의 모집단 모형의 EMS를 완성하고, F 통계량을 구하여라. 단 수준 수는 4, 5이며, 반복수는 2이다.

	4	5	2		
	R	F	R		F 통계량
요인	i	j	k	EMS	
$\alpha_i$	1	5	2	$\sigma^2 + 10\sigma_\alpha^2$	$MS_A / MSE$
$\beta_j$	4	0	2	$\sigma^2 + 2\sigma_{\alpha\beta}^2 + 8 \sum_{j=1}^5 \beta_j^2 / (b-1)$	$MS_B / MS_{AB}$
$\alpha\beta_{ij}$	1	0	2	$\sigma^2 + 2\sigma_{\alpha\beta}^2$	$MS_{AB} / MSE$
$\epsilon_{ijk}$	1	1	1	$\sigma^2$	

[문제3] 요인모형  $Y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \epsilon_{ijk}$   $i = 1, \dots, a$   $j = 1, \dots, b$   $k = 1, \dots, n$  에 대하여  
아래 오른쪽 분지모형의 ANOVA 표를 완성하여라. (제공합 및 자유도)

요인모형(고정)		
요인	제공합	자유도
A	$SS_A$	$a - 1$
B	$SS_B$	$b - 1$
AB	$SS_{AB}$	$(a - 1)(b - 1)$
잔차	$SSE$	$ab(n - 1)$
총합	$SST$	$abn - 1$

2단계분지모형(고정)		
요인	제공합	자유도
A	$SS_A = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{y}_{i..} - \bar{y}_{...})^2$	$a - 1$
B	$SS_{B(A)} = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{y}_{ij.} - \bar{y}_{i..})^2$	$a(b - 1)$
잔차	$SSE = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (y_{ijk} - \bar{y}_{ij.})^2$	$ab(n - 1)$
총합	$SST = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (y_{ijk} - \bar{y}_{...})^2$	$abn - 1$

[문제5] 아래의 BIB 모형에 대하여 DO-END loop를 사용하여 자료입력을 하여라.  
(DATA; INPUT additive car mileage @@; CARDS; 1 2 17 1 3 14 .....)

car / additive	1	2	3	4	5
1	.	14	12	13	11
2	11	14	.	12	12
3	14	.	13	11	10
4	13	13	12	12	.
5	12	10	9	.	8

```
DATA;
INPUT additive car mileage @@;
CARDS;
1 2 11 1 3 14 1 4 13 1 5 12
2 1 14 2 2 14 2 4 13 2 5 10
3 1 12 3 3 13 3 4 12 3 5 9
4 1 13 4 2 12 4 3 11 4 4 12
5 1 11 5 2 12 5 3 10 5 5 8
;
```

[문제6]

- Annalysis of Covariance 이란 무엇인가?

분산분석과 회귀분석의 특성을 결합한 분석기법 => 공변수의 효과를 보정한 분산분석의 기법

- 공변수(covariate)란?

반응 변수 Y와 더불어 선형관계의 또 다른 변수 X가 실험계획 설계자가 조절할 수 없고, 반응 변수 Y와 함께 관측되는 변수.

- 반응변수(time), 분류변수(store), 공변수(price) 로 주어졌을 때

```
PROC GLM;
CLASS store;
MODEL time = store price / SOLUTION;
```



[문제7] 아래 문항에 대하여 간단히 답하여라.

– 측면도 plot 의 용도는?

주요상호작용이 어느 수준에서 구체적으로 크고 작은지 확인하기 위함이다. 요인내의 수준에 따라서 각각의 그래프를 그려보면 평행일 때 상호작용이 존재하는지 않는 것을 알 수 있다. 따라서 그래프를 통해서 상호작용 효과가 있는지를 알 수 있다.

– 분산요소 추정의 필요성은?

랜덤효과일 때의 분산분석의 관심은 처리 수준 간의 분산이 존재하는지를 파악하는 것이다. 즉 관측값의 총분산은 처리 간의 분산을 측정하는 요소와 처리 내의 분산을 측정하는 요소로 분해되는데 이를 통해 각 요인의 효과가 존재하는 지 검정할 수 있기 때문이다.

– 고정효과와 랜덤효과를 비교할 때, 어느 효과가 가설검정을 위한 통계량 값이 작아지는 경향이 있나? 통계량 값이 작아지는 경향이란 무엇을 의미하나?

## 2015년 실험계획법 기말고사

[문제2] 아래의 문항에 대하여 간단히 답하여라.

- 고정효과와 랜덤효과의 결과해석에 따른 주된 차이점은?

고정효과의 관심은 특정개별 처리 효과간의 차이가 존재하는지를 검정하는 것이고, 랜덤효과의 관심은 처리 수준 간의 분산이 존재하는지를 파악하는 것이다.

[문제3] A, B, C 요인이 모두 랜덤(R)인  $Y_{ijkl} = \mu + \alpha_i + \beta_j + \gamma_k + \alpha\beta_{ij} + \alpha\gamma_{ik} + \beta\gamma_{jk} + \alpha\beta\gamma_{ijk} + \epsilon_{ijkl}$  의

삼요인계획법 모형의 EMS를 완성하고, 각 효과검정을 위한 F 통계량을 표현하여라.

(단, 세요인 수준 수는 2, 3, 4 이며, 반복수는 2 이다.)

요인	2 R i	3 R j	4 R k	2 R l	F 통계량
$\alpha_i$	1	3	4	2	$MS_A / MSE$
$\beta_j$	2	1	4	2	$MS_B / MSE$
$\gamma_k$	2	3	1	2	$MS_C / MSE$
$\alpha\beta_{ij}$	1	1	4	2	$MS_{AB} / MSE$
$\alpha\gamma_{ik}$	1	3	1	2	$MS_{AC} / MSE$
$\beta\gamma_{jk}$	2	1	1	2	$MS_{BC} / MSE$
$\alpha\beta\gamma_{ijk}$	1	1	1	2	$MS_{ABC} / MSE$
$\epsilon_{(ijk)l}$	1	1	1	1	

[문제4] 라틴방격모형  $Y_{ijk} = \mu + \alpha_i + \beta_j + \gamma_k + \epsilon_{ijk} \quad i, j, k = 1, \dots, p$  (고정효과)의 분산분석표 빈칸을 완성하여라.

요인	제곱합	자유도	평균제곱합	F 통계량
처리	$SS_t = p \sum_{k=1}^p (\bar{y}_{..k} - \bar{y}_{...})^2$	$p-1$	$MS_t = \frac{SS_t}{p-1}$	$\frac{MS_t}{MSE}$
열블럭	$SS_C = p \sum_{i=1}^p (\bar{y}_{i..} - \bar{y}_{...})^2$	$p-1$	$MS_C = \frac{SS_C}{p-1}$	.
행블럭	$SS_R = p \sum_{j=1}^p (\bar{y}_{.j.} - \bar{y}_{...})^2$	$p-1$	$MS_R = \frac{SS_R}{p-1}$	.
잔차	$SS_t = p \sum_{i=1}^p \sum_{j=1}^p (\bar{y}_{ijk} - \bar{y}_{i..} - \bar{y}_{.j.} - \bar{y}_{..k} + 2\bar{y}_{...})^2$	$(p-1)(p-2)$	$MSE = \frac{SSE}{(p-1)(p-2)}$	.
총합	$SS_t = p \sum_{k=1}^p (\bar{y}_{..k} - \bar{y}_{...})^2$	$p^2-1$		

[문제5] 2단계 분지모형  $Y_{ijk}, \quad i = 1, \dots, a \quad j = 1, \dots, b \quad k = 1, \dots, n$  (고정효과)의 분산분석표(B요인이 A요인에 포함) 빈칸을 완성하여라.

요인	제곱합	DF	MS	F 통계량
A	$SS_A = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{y}_{i..} - \bar{y}_{...})^2$	$a-1$	$MS_A = \frac{SS_A}{a-1}$	$\frac{MS_A}{MSE}$
B(A)	$SS_{B(A)} = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (\bar{y}_{ij.} - \bar{y}_{i..})^2$	$a(b-1)$	$MS_{B(A)} = \frac{SS_{B(A)}}{a(b-1)}$	$\frac{MS_{B(A)}}{MSE}$
잔차	$SSE = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (y_{ijk} - \bar{y}_{ij.})^2$	$ab(n-1)$	$MSE = \frac{SSE}{ab(n-1)}$	.
총합	$SST = \sum_{i=1}^a \sum_{j=1}^b \sum_{k=1}^n (y_{ijk} - \bar{y}_{...})^2$	$abn-1$		

[문제6] 의료보험제도의 여론(opinion)을 조사하고자 3단계 분지모형을 설계하였다. 전국을 두 지역(area : 중부, 남부)으로 나누고, 각 지역으로부터 도(state) 및 도시(city)를 선택하였다. 아래의 결과를 이용하여 ① area, state, city 세요인의 고정(F) 혹은 랜덤(R) 여부를 결정한 후에 명령문을 완성하여야.

The GLM Procedure					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	11	8996.45833	817.76894	4.58	0.0071
Error	12	2141.50000	178.45833		
Corrected Total	23	11136.95833			

Source	DF	Type III SS	Mean Square	F Value	Pr > F
area	1	315.375000	315.375000	1.77	0.2084
state(area)	4	6453.833333	1613.458333	9.04	0.0013
city(area*state)	6	2226.250000	371.041667	2.08	0.1322

Type III Expected Mean Square  
area:  $\text{Var(Error)} + 2 \cdot \text{Var(city(area*state))} + Q(\text{area})$   
state(area):  $\text{Var(Error)} + 2 \cdot \text{Var(city(area*state))} + Q(\text{state(area)})$   
city(area\*state):  $\text{Var(Error)} + 2 \cdot \text{Var(city(area*state))}$

Tests of Hypotheses Using the Type III MS for city(area\*state) as an Error Term

Source	DF	Type III SS	Mean Square	F Value	Pr > F
area	1	315.375000	315.375000	(2)	0.3921
state(area)	4	6453.833333	1613.458333	(3)	0.0545

Bonferroni Test for opinion

Bon Grouping	Mean	N	area
A	56.667	12	1
A	49.417	12	2

Type I Estimates

Variance Component	Estimate
Var(city(area*state))	(4) = 96.29167
Var(Error)	(5) = 178.45833

② PROC GLM;

```
CLASS area state city;
MODEL opinion = area state(area) city(state area);
RANDOM city(area*state);
TEST H=area state(area) E=city(state area);
MEASNS area / BON E=city(state area);
```

③ PROC VARCOMP;

```
CLASS area state city;
MODEL opinion = area state(area) city(state area) / FIXED=2;
```

- PROC NESTED 프로시저를 적용하는 경우는?

분지모형을 위한 명령문으로 요인이 모두 랜덤인 경우를 가정한다.

- 위 결과의 밑줄 친 5 빈칸 (1) - (5) 을 서술하거나, 계산 값을 추정하여라.

- (1) city(area\*state)
- (2) 0.85
- (3) 4.35
- (4) 96.29167
- (5) 178.45833

- 만일 유의수준  $\alpha = 0.05$  하에서 area 요인효과가 존재하는가? 이유는(혹은 p-값은)?

유의수준 5% 하에서 area는 p-값이 0.3921로 "요인효과가 존재하지 않는다."라는 귀무가설을 채택한다. 요인효과는 존재하지 않는다.

- 만일 유의수준  $\alpha = 0.05$  하에서 state 요인효과가 존재하는가? 이유는(혹은 p-값은)?

유의수준 5% 하에서 area는 p-값이 0.0545로 "요인효과가 존재하지 않는다."라는 귀무가설을 채택한다. 요인효과는 존재하지 않는다.

- 만일 유의수준  $\alpha = 0.05$  하에서 city 요인효과가 존재하는가? 이유는(혹은 p-값은)?

유의수준 5% 하에서 area는 p-값이 0.1322로 “요인효과가 존재하지 않는다.”라는 귀무가설을 채택한다. 요인효과는 존재하지 않는다.

- 두 지역( area : 중부, 남부) 중에서 평균수치가 높은 지역은?

area1은 평균이 56.667, area2는 평균이 49.417 이지만 BON 검정 결과 차이가 없다.

- 만일 3요인 모두를 랜덤(R)으로 가정한 경우의 PROC NESTED 명령문은?

```
PROC NESTED;  
CLASS area state city;  
VAR opinion;
```

## 2014년 실험계획법 기말고사

[문제1] EMS 열이 다음과 같이 주어진 3요인 모형을 가정하자.

Source	EMS
A	$\text{Var}(\text{Error}) + \text{Var}(\text{ABC}) + 6 \text{Var}(\text{AB}) + 24 \text{Var}(\text{A})$
B	$\text{Var}(\text{Error}) + \text{Var}(\text{ABC}) + 6 \text{Var}(\text{AB}) + \text{Q}(\text{B})$
C	$\text{Var}(\text{Error}) + \text{Var}(\text{ABC}) + 8 \text{Var}(\text{AC}) + \text{Q}(\text{C})$
AB	$\text{Var}(\text{Error}) + \text{Var}(\text{ABC}) + 6 \text{Var}(\text{AB})$
AC	$\text{Var}(\text{Error}) + \text{Var}(\text{ABC}) + 8 \text{Var}(\text{AC})$
BC	$\text{Var}(\text{Error}) + \text{Var}(\text{ABC}) + 8 \text{Q}(\text{BC})$
ABC	$\text{Var}(\text{Error}) + \text{Var}(\text{ABC})$
Error	$\text{Var}(\text{Error})$

- 주요인 A, B, C 의 Fixed 및 Random 인지를 밝혀라. 그 이유는?

분산요소의 오른쪽 마지막의 제곱항(상수)이 존재하지 않으므로 A는 Random 요인이다.

분산요소의 오른쪽 마지막의 제곱항(상수)이 Q(B) 존재하므로 B는 Fixed 요인이다.

분산요소의 오른쪽 마지막의 제곱항(상수)이 Q(C) 존재하므로 C는 Fixed 요인이다.

- 주효과 및 상호작용을 포함한 각 요인들의 효과유무에 따른 F 통계량을 표현하여라.

요인 A : $MS_A / MSE$	요인 AC : $MS_{AC} / MSE$
요인 B : $MS_B / MS_{AB}$	요인 BC : $MS_{BC} / MS_{ABC}$
요인 C : $MS_C / MS_{AC}$	요인 ABC : $MS_{ABC} / MSE$
요인 AB : $MS_{AB} / MSE$	

[문제2] 다음과 같이 SAS 결과가 주어진 2요인 모형을 가정하자.

```

PROC GLM;
  CLASS M (a) T ;
  MODEL M (b) T M*T;
  RANDOM T (c) M*T;
  TEST T (d) M*T;

  PROC VARCOMP method=type1;
    model ;
  ;

```

Dependent Variable: sales						
Source	DF	Type III SS	Mean Square	F Value	Pr > F	
MATERIAL	2	10683.72222	5341.86111	7.91	0.0020	
TEMP	2	39118.72222	19559.36111	28.97	0.0001	
MATERIAL*TEMP	4	9613.77778	2403.44444	3.56	0.0186	
Error	27	18230.75000	675.21296			
Corrected Total	35	77646.97222				

Source	Type III Expected Mean Square
MATERIAL	$\text{Var}(\text{Error}) + 4 \text{Var}(\text{MATERIAL*TEMP}) + \text{Q}(\text{MATERIAL})$
TEMP	$\text{Var}(\text{Error}) + 4 \text{Var}(\text{MATERIAL*TEMP}) + 12 \text{Var}(\text{TEMP})$
MATERIAL*TEMP	$\text{Var}(\text{Error}) + 4 \text{Var}(\text{MATERIAL*TEMP})$

Tests of Hypotheses using the Type III MS for M*T as an error term						
Source	DF	Type III SS	Mean Square	F Value	Pr > F	
MATERIAL	2	10683.72222	5341.86111	(h)	0.2243	
TEMP	2	39118.72222	19559.36111	(i)	0.0389	

Variance Component	Estimate
Var(TEMP)	1429.7
Var(MATERIAL*TEMP)	(j)
Var(Error)	(k)

$\sigma^2_{AB} = \frac{MS_A - MS_{AB}}{n}$   
 $\hat{\sigma}^2 = 675.21296$

- 변수를 이용하여 실험계획의 관심 문제를 해석하여라.

MATERIAL과 TEMP 각 수준간의 요인효과 및 MATERIAL\*TEMP의 차이가 존재하는지 알고자 검증하는 것.

- 주요인들의 Fixed 및 Random 여부를 밝혀라. 그 이유는?

분산요소의 오른쪽 마지막의 제곱항(상수)이 Q(MATERIAL) 존재하므로 MATERIAL은 Fixed 요인이다.  
분산요소의 오른쪽 마지막의 제곱항(상수)이 존재하지 않으므로 TEMP는 Random 요인이다.

- 주효과 및 상호작용의 효과유무를 위한 F 통계량을 표현하여라.

MATERIAL :  $MS_A / MS_{AB}$   
TEMP :  $MS_B / MSE$   
MATERIAL\*TEMP :  $MS_{AB} / MSE$

- 만일 유의수준  $\alpha = 0.05$  하에서 MATERIAL 요인효과가 존재하는가? 이유는(혹은 p-값은)?

유의수준 5% 하에서 MATERIAL은 p-값이 0.3921로 "요인효과가 존재하지 않는다."라는 귀무가설을 채택한다. 요인효과는 존재하지 않는다.

- 만일 유의수준  $\alpha = 0.05$  하에서 TEMP 요인효과가 존재하는가? 이유는(혹은 p-값은)?

유의수준 5% 하에서 TEMP는 p-값이 0.0389로 "요인효과가 존재하지 않는다."라는 귀무가설을 기각한다. 요인효과는 존재한다.

- 만일 유의수준  $\alpha = 0.05$  하에서 상호작용의 존재여부는? 이유는(혹은 p-값은)?

유의수준 5% 하에서 상호작용은 p-값이 0.0186로 "상호작용이 존재하지 않는다."라는 귀무가설을 기각한다. 상호작용은 존재한다.

- (a) - (f) 의 PROC GLM 명령문을 완성하여라.

```
PROC GLM;  
CLASS MATERIAL TEMP;  
MODEL sales = MATERIAL TEMP MATERIAL*TEMP;  
RANDOM TEMP MATERIAL*TEMP / TEST;  
TEST H=MATERIAL TEMP E=MATERIAL*TEMP;
```

- (g) 및 (h) - (k)의 값은?

(g) MATERIAL\*TEMP  
(h)  $5341/2403 = 2.22$   
(i)  $19559/675 = 28.98$   
(j)  $2403.4444-675.21296 / 4 = 432.06$   
(k) 675.21296

- 만약 2요인 material 및 temp 모두가 랜덤인, 랜덤모형인 경우의 PROC GLM과 PROC VARCOMP 명문은?

```
PROC GLM;  
CLASS MATERIAL TEMP;  
MODEL sales = MATERIAL TEMP MATERIAL*TEMP;  
RANDOM MATERIAL TEMP MATERIAL*TEMP / TEST;  
TEST H = MATERIAL TEMP E = MATERIAL*TEMP;  
  
PROC VARCOMP;  
CLASS MATERIAL TEMP;  
MODEL sales = MATERIAL TEMP MATERIAL*TEMP / FIXED=0;
```